

APPLICATION

FOR

UNITED STATES LETTERS PATENT

**TITLE: DETERMINING TARGET OPERATING
 FREQUENCIES FOR A MULTIPROCESSOR SYSTEM**

INVENTORS: SHIVNANDAN D. KAUSHIK; PING SAGER

Express Mail No. EV 337934698 US

Date: July 23, 2003

DETERMINING TARGET OPERATING FREQUENCIES FOR A
MULTIPROCESSOR SYSTEM

Background

The present invention relates to power management of
5 a computer system and more particularly to power
management for a multiprocessor system.

Multiprocessor computer systems include multi-
threaded processors in which a single physical processor
is segmented into multiple logical processors, and
10 multicore processors in which multiple processor cores
are present in a single package or plural packages.

Mobile computers such as notebook personal computers
(PCs) typically incorporate certain power management
techniques. One such technique is an adaptive technology
15 which provides for changing both the operating voltage
and frequency of the processor such that transition to a
lower frequency (i.e., performance) point leads to a
cubic reduction in power consumption by the processor at
the lower frequency point. Current operating systems use
20 this technique so that the processor is placed at an
operating frequency that matches the processor
utilization. For example if the processor is idle 50% of
the time, then the operating system places the processor
at a frequency that is 50% of the maximum operating
25 frequency. However, such techniques can have a

deleterious effect on multiprocessor systems, as a frequency change on one processor might get applied to other system processors, which may due to their processor utilization desire higher operating frequencies. Thus a need exists to provide power management techniques for multiprocessor systems.

Brief Description of the Drawings

FIG. 1 is flow diagram of a method in accordance with one embodiment of the present invention.

FIG. 2 is a flow diagram of determining a target frequency for a physical processor package in accordance with one embodiment of the present invention.

FIG. 3 is a block diagram of a system with which embodiments of the present invention may be used.

Detailed Description

Referring to FIG. 1, shown is flow diagram of a method in accordance with one embodiment of the present invention. As shown in FIG. 1, the method begins by obtaining parameter information for each processor of a system (block 110). In some embodiments, the processors may be logical processors while in other embodiments the processors may be multicore processors, either in a single physical package or separate physical packages. Further, in certain

embodiments the processors may have dependencies with respect to power utilization and/or frequency transitions.

While parameter information may vary in different embodiments, in one embodiment the parameter information
5 may include a sample interval (S), an up threshold percentage (U), an up threshold time (UT), a down threshold percentage (D), and a down threshold time (DT), as will be discussed below.

Based upon the parameter information, a utilization
10 value may be determined for each processor (block 120). In one embodiment, the utilization value may be a transition decision for each processor. For example, in one embodiment an up/down decision may be made, meaning that a decision may be made as to whether that processor frequency
15 should be raised or lowered. In alternate embodiments, a utilization value such as processor usage as a percentage of a maximum operating frequency may be used to obtain an up/down decision. Still further, in other embodiments a utilization value may be determined without reference to
20 transition decisions.

Finally in the embodiment of FIG. 1, based upon the utilization values, a target processor frequency may be identified (block 130). For example, in one embodiment up/down decisions for the processors may be used to
25 determine a target processor frequency for the multiple processors.

Thus in certain embodiments, the present invention may provide an adaptive algorithm for coordinating between logical processors of a multithreaded processor or between multiple cores of a multicore processor. In certain
5 embodiments, the multithreaded processor may be enabled for Hyper-Threading technology. In certain multicore processors, frequency control logic may be shared among the cores.

In one embodiment of the present invention, an
10 algorithm may use a number of parameters to determine an up/down decision for each processor present. These parameters may include a sample interval (S), an up threshold percentage (U), an up threshold time (UT), a down threshold percentage (D), and a down threshold time (DT),
15 in one embodiment.

In such an embodiment, the sample interval may be the time interval at which frequency determinations may be made. While the time for such intervals may vary, in certain embodiments, S may be between approximately 10 and
20 100 milliseconds (ms). The up threshold percentage may be a processor utilization percentage above which an up decision may be made. In one embodiment, U may be approximately 60%, meaning that if a processor usage is more than 60%, the transition decision may be an up
25 decision. The up threshold time may be the interval of time at which the up threshold percentage is analyzed. In

one embodiment, UT may be approximately 100 ms. The down threshold percentage may be a processor utilization percentage below which a down decision is made. In one embodiment, DT may be approximately 30%. The down
5 threshold time may be the time interval at which down transition decisions may be made. In certain embodiments, DT may be between approximately 500 to 1000 ms.

In one embodiment, an algorithm may include two phases applied at every S interval of time, which in certain
10 embodiments may occur whenever a given processor (e.g., processor X) enters an operating system (OS) idle loop. In the first phase, an up/down decision may be determined for every processor of the system (including itself). In certain embodiments, the decision may be determined for all
15 processors with which processor X shares transition logic and/or has transition dependencies.

In this embodiment, the following algorithm may be applied for each processor: if processor utilization for processor X over the past UT time interval is more than U%,
20 then the transition decision for processor X may be up. However, if processor utilization for processor X over the past DT interval of time is less than D%, then the transition decision for processor X may be down. This algorithm may be performed for each processor of the
25 multiprocessor system.

Next, in this embodiment the second phase may determine a collective decision for all processors of the system. In such manner, the appropriate target frequency to which the physical package should transition may be identified. While many options may be used for the collective decision, in certain embodiments the decision may depend on whether there is a preference for performance optimization or power saving.

In one embodiment, the following option may provide a preference for performance by responding immediately to a need for compute power by any processor in a system by transitioning the physical package to a higher frequency. Specifically, if any processor has reached an up transition decision, then the physical package may be transitioned to a higher frequency. The target higher frequency for the package may be the frequency operating point closest to the maximum utilization among all the processors multiplied by the maximum operating frequency, in one embodiment. On the other hand, if all processors have reached a down transition decision, then the physical package may be transitioned to a lower frequency. The target lower frequency for the package may be the frequency operating point closest to the maximum utilization among all processors multiplied by the maximum operating frequency, in one embodiment.

Referring now to FIG. 2, shown is a flow diagram of a method for determining a target frequency for a physical package in accordance with one embodiment of the present invention. As shown in FIG. 2, up/down decisions may be
5 obtained for each processor (block 210). Next it may be determined whether there is an up decision for any of the processors (diamond 215). If so, a target frequency may be selected based on the highest utilization processor (block 220). For example, in one embodiment the target frequency
10 may be equal to a frequency operating point of the processor closest to the maximum utilization among all of the processors multiplied by the maximum operating frequency. Then the processor package may be transitioned to the new higher frequency (block 225).

15 If instead at diamond 215 it is determined that there is no up decision for any processor, it next may be determined whether all processors have a down decision (diamond 230). If all such processors have a down decision, then a target frequency may be selected based on
20 the highest utilization processor (block 240). For example, in one embodiment a target frequency may be equal to a frequency operating point closest to the maximum utilization among all of the processors multiplied by the maximum operating frequency. Then, the processor package
25 may be transitioned to the new lower target frequency (block 250).

Still referring to FIG. 2, if all of the processors did not have a down decision, the target frequency of the processor package may remain at its present frequency (block 260).

5 Other embodiments may choose to combine the individual up or down decisions for each processor in a different fashion, leading to a different target frequency and power for the processor package. For instance, it may be possible to combine the individual up or down decisions for
10 each processor such that the processor package makes a down transition if any processor has a down decision and an up transition only if all the processors have an up decision. In certain embodiments, an operating system may choose to use the different embodiments at different times during its
15 operation depending on the specific power/performance needs at that point of time.

Embodiments may be implemented in a computer program. As such, these embodiments may be stored on a storage medium having stored thereon instructions which can be used
20 to program a computer system such as a notebook computer, wireless device or the like to perform the embodiments. The storage medium may include, but is not limited to, any type of disk including floppy disks, optical disks, compact disk read-only memories (CD-ROMs), compact disk rewritables (CD-
25 RWs), and magneto-optical disks, semiconductor devices such as read-only memories (ROMs), random access memories (RAMs)

such as dynamic and static ROMs, erasable programmable read-only memories (EPROMs), electrically erasable programmable read-only memories (EEPROMs), flash memories, magnetic or optical cards, or any type of media suitable
5 for storing electronic instructions. Similarly, embodiments may be implemented as software modules executed by a programmable control device, such as a computer processor or a custom designed state machine.

FIG. 3 is a block diagram of a representative data
10 processing system, namely computer system 300 with which embodiments of the invention may be used. In one embodiment, computer system 300 includes a processor 310, which may include a general-purpose or special-purpose processor such as a microprocessor, microcontroller,
15 application specific integrated circuit (ASIC), a programmable gate array (PGA), and the like. As shown in FIG. 3, processor 310 may be a multithread or multicore processor including a first processor 303 (P1) and a second processor 307 (P2).

20 The processor 310 may be coupled over a host bus 315 to a memory hub 330 in one embodiment, which may be coupled to a system memory 320 via a memory bus 325. The memory hub 330 may also be coupled over an Advanced Graphics Port (AGP) bus 333 to a video controller 335, which may be
25 coupled to a display 337. The AGP bus 333 may conform to the Accelerated Graphics Port Interface Specification,

Revision 2.0, published May 4, 1998, by Intel Corporation, Santa Clara, California.

The memory hub 330 may also be coupled (via a hub link 338) to an input/output (I/O) hub 340 that is coupled to a
5 input/output (I/O) expansion bus 342 and a Peripheral Component Interconnect (PCI) bus 344, as defined by the PCI Local Bus Specification, Production Version, Revision 2.1 dated in June 1995, or alternately a bus such as the PCI Express bus, or another third generation I/O interconnect
10 bus. The I/O expansion bus 342 may be coupled to an I/O controller 346 that controls access to one or more I/O devices. As shown in FIG. 3, these devices may include in one embodiment storage devices, such as a floppy disk drive 350 and input devices, such as keyboard 352 and mouse 354.
15 The I/O hub 340 may also be coupled to, for example, a hard disk drive 356 as shown in FIG. 3. It is to be understood that other storage media may also be included in the system. In an alternate embodiment, the I/O controller 346 may be integrated into the I/O hub 340, as may other
20 control functions.

The PCI bus 344 may be coupled to various components including, for example, a flash memory 360. Further shown in FIG. 3 is a wireless interface 362 coupled to the PCI bus 344, which may be used in certain embodiments to
25 communicate with remote devices. As shown in FIG. 3, wireless interface 362 may include a dipole or other

antenna 363 (along with other components not shown in FIG. 3). In various embodiments, wireless interface 362 may be coupled to system 300, which may be a notebook personal computer, via an external add-in card, or an embedded
5 device. In other embodiments, wireless interface 362 may be fully integrated into a chipset of system 300.

Although the description makes reference to specific components of the system 300, it is contemplated that numerous modifications and variations of the described and
10 illustrated embodiments may be possible.

While the present invention has been described with respect to a limited number of embodiments, those skilled in the art will appreciate numerous modifications and variations therefrom. It is intended that the appended
15 claims cover all such modifications and variations as fall within the true spirit and scope of this present invention.